

What is claimed is:

1. A method of modifying data comprising: loading a computer system including a processor and a display device with a computer-executable program comprising a software module and a user interface having a representation of available transformations, a sequence assembly area, and a plurality of user-selectable, user-sequenciable operations; choosing any number of said operations for application to said data; assembling and optionally displaying the chosen operations in said sequence assembly area; and applying the chosen sequence of operations to said data to produce modified data for storage or display.

2. The method of claim 1 further comprising selecting microarray data as said data.

3. The method of claim 1 or claim 2 wherein each of said operations includes an associated visual representation and performs a specific operation on said data, and wherein each of said operations may include an associated dialog box prompting a user to choose one or more data preparation parameters, and said software module permits a user to drag one or more of said visual representations from said representation of available transformation into said sequence assembly area.

4. The method of claim 1 further comprising selecting said data from microarray data, and arranging said data with a graphical user interface, data set builder, that includes a data source list from which a user can define relationships or associations of the data including pairs of data sources and replicated data sources, as desired.

5. The method of claim 4 further comprising providing said data set builder with the capacity to prepare a data set that includes single data sources, paired data sources, or replicated data sources, at a user's option.

6. The method of claim 1 or claim 2 further comprising choosing said operations from the group consisting of background correction of data values, omission of one or more data item based on a characteristic value, combining replicate data, addition of one or more missing data, modification of data values to raise those below a specified threshold value to the specified threshold value, transforming data, combining replicated data, forming a ratio of two or more data,

taking the difference between data, omitting data values based on its value, and normalizing data.

7. The method of claim 6 further comprising choosing said normalizing operation, said normalizing operation including the steps of dividing data values into groups of neighboring values, and determining and applying a specific normalizing factor for each said group.

8. The method of claim 7 further comprising the step of specifying the size of each said group to ensure that a predetermined number of values are in said group

9. A system for modifying data comprising a memory storing said data, a processor for accessing said data from said memory, and optionally a display for displaying said data, said system also including a software module and a user interface having a representation of available transformations, a sequence assembly area, and a plurality of user-selectable, user-sequenciable operations, said software module permitting a user to choose any number of said operations for application of said data, to assemble the chosen operations in said sequence assembly area, and to apply the chosen sequence of operations to said data to produce modified data.

10. The system of claim 9 wherein said data is microarray data.

11. The system of claim 9 or claim 10 wherein each of said operations includes an associated visual representation and performs a specific operation on said data, and wherein each of said operations may include an associated dialog box prompting a user to choose one or more available data preparation parameters, and said software module permits a user to drag one or more of said visual representation from said representation of available transformation into said sequence assembly area.

12. The system of claim 9 further comprising a data set builder module that includes a data source list from which a user can define structures of the data including pairs of data sources and replicated data sources, as desired

13. The system of claim 12 wherein said data set builder has the capacity to prepare a data set from single data sources, from paired data sources, or from replicated data sources, at a user's option.

14. The system of claim 9 or claim 10 wherein said operations are selected from the group consisting of background correction of data, omission of one or more desired data, combining replicate data, addition of one or more missing data at a user's option, modification of data values to raise those below a specified threshold value to the specified threshold value, transforming data to the log of the data, combining replicated data, forming a ratio of two or more data, taking the difference between data, omitting data values below a specified threshold value, and normalizing data.

15. A computer readable medium including a computer-executable program comprising a user interface having a representation of available transformations, a sequence assembly area, and a plurality of user-selectable, user-sequentially operations, said medium having stored thereon one or more sequences of instructions for mathematically modifying data, said one or more sequences of instructions causing one or more processors to perform a plurality of acts, said acts comprising: choosing any number of said operations for application to said data; assembling and optionally displaying the chosen operations in said sequence assembly area; and applying the chosen sequence of operations to said data to produce modified data.

16. The computer readable medium of claim 15 wherein each of said operations includes an associated visual representation and performs a specific operation on said data, and wherein each of said operations may include an associated dialog box prompting a user to choose one or more available data preparation parameters, and said software module permits a user to drag one or more of said visual representation from said representation of available transformation into said sequence assembly area.

17. The computer readable medium of claim 15 or claim 16 wherein said operations are selected from the group consisting of background correction of data values, omission of one or more desired data, combining replicate data, addition of one or more missing data at a user's option, modification of data values to raise those below a specified threshold value to the specified threshold value, transforming data to the log of the data, combining replicated data, forming a ratio of two or more data, taking the difference between data, omitting data values below a specified threshold value, and normalizing data.

18. A method for normalizing data comprising the steps of: dividing the

data into a plurality of groups, wherein the number of groups is a function of the range and number of values and for calculating a normalization correction for each group.

19. The method of claim 18 wherein the normalized values in said groups are determined such that a particular distribution (such as the scatterplot of the values measured on different channels) is brought to a desired shape.

20. The method of claim 18 or 19 where the groups overlap to such a degree that the computation is efficient in terms of the number of operations executed for a given data set.

21. The method of claim 18, 19 or 20 wherein the desired shape is a line of approximately slope 1.

22. The method of claim 18, 19, 20 or 21 wherein there are no large discontinuities between adjacent groups.

09981865-101701